Data Glacier

Your Deep Learning Partner

Bank marketing (campaign)

Presented to ABC bank

Group name: Fleifel-solo
Name: Rania Tarek Fleifel
Email: raniatarekfleifel@gmail.com
Country: Egypt, **College:** Cairo university Faculty of engineering
Specialization: Data science
Internship Batch: LISUM13: 30

Submission date: 14th December 2022
Github repo: https://github.com/RaniaFleifel/Data-glacier-internship/tree/main/Data%20science%20project_Bank%20Marketing%20Campaign

Bank marketing (campaign)

Business Problem:
The ultimate goal of the partnership with ABC bank is to develop a predictive model that enables the bank to target costumers who’re more probable to invest in their new term deposit product.

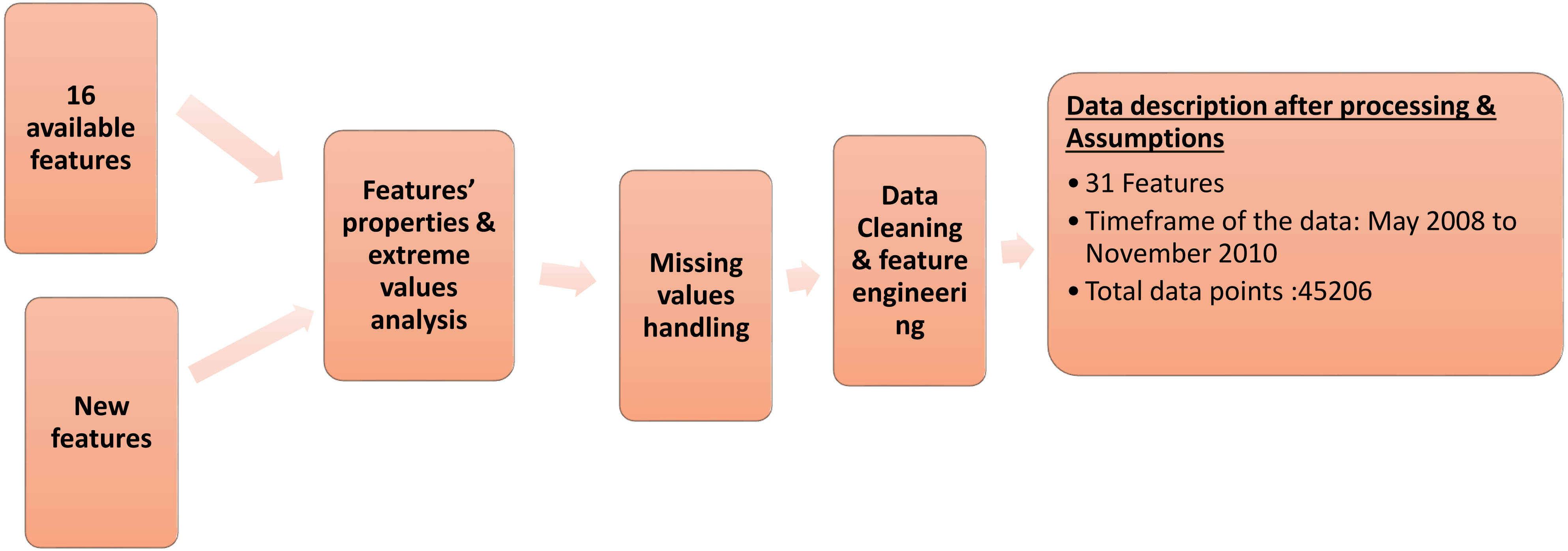
Objective :

- Translate data into insights to help ABC markting team target costumers who are most probable to invest in the new deposit product. This presentation aims to show the exploratory data analysis efforts done on the data to detangle how different information relate. This information will help ABC bank tailor the product to their target audience.
- Tenor, interest rate and payment frequency terms of the product can be tuned according to a costumer’s personal and social standing.

Exploratory analysis steps:

1) Data preparation and wrangling
2) Data Analysis
3) Proposed modeling technique

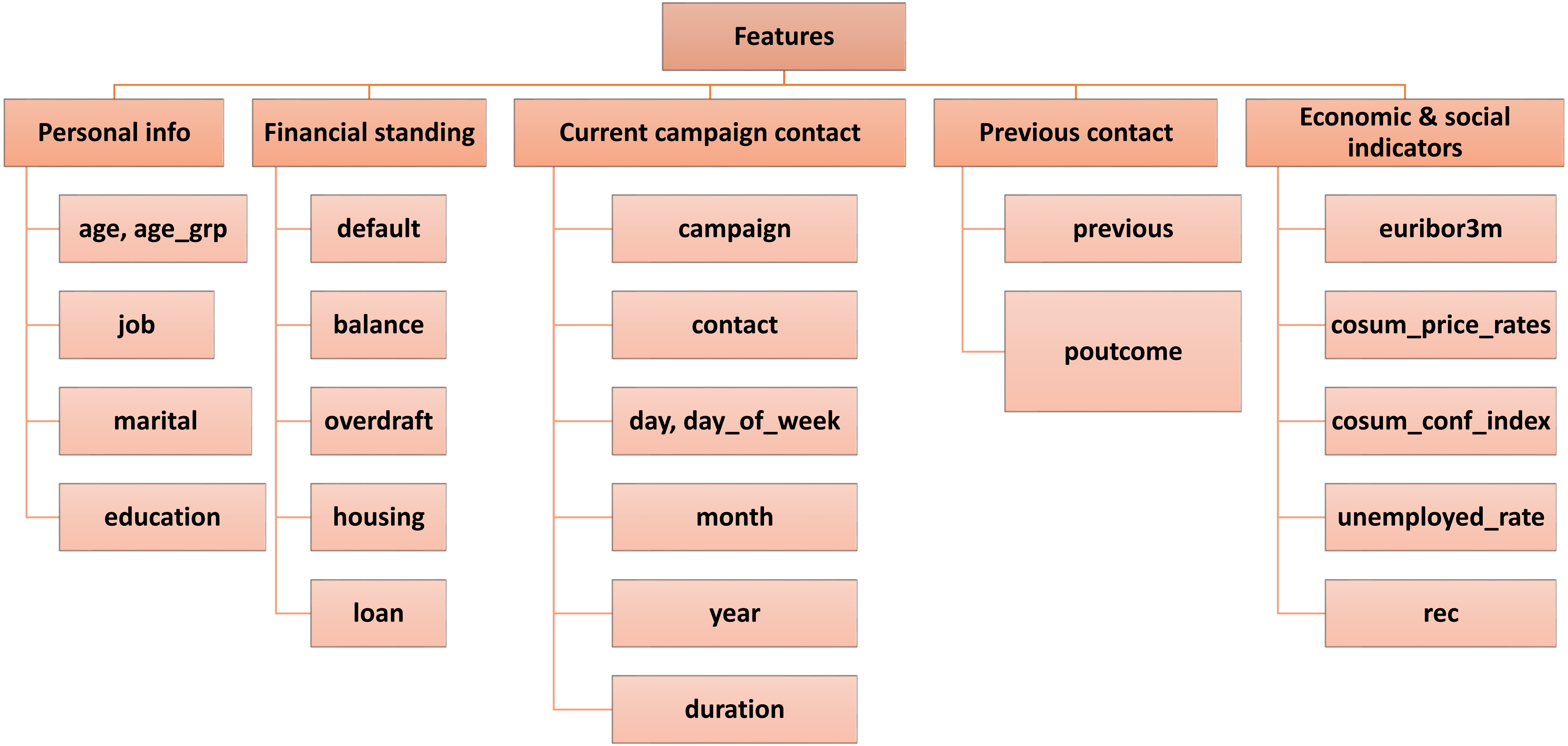
Data preparation and wrangling



Data Assumptions:

- Data provided is correct, there's no wrong reporting of data .
- The data spans all year including weekends and holidays.
- Means of communication, other than phone and mobile, are not considered.
- For Features with "unknown"/missing data , regardless how they're handled, induce a binary indicator to indicate whether this data was present or not.

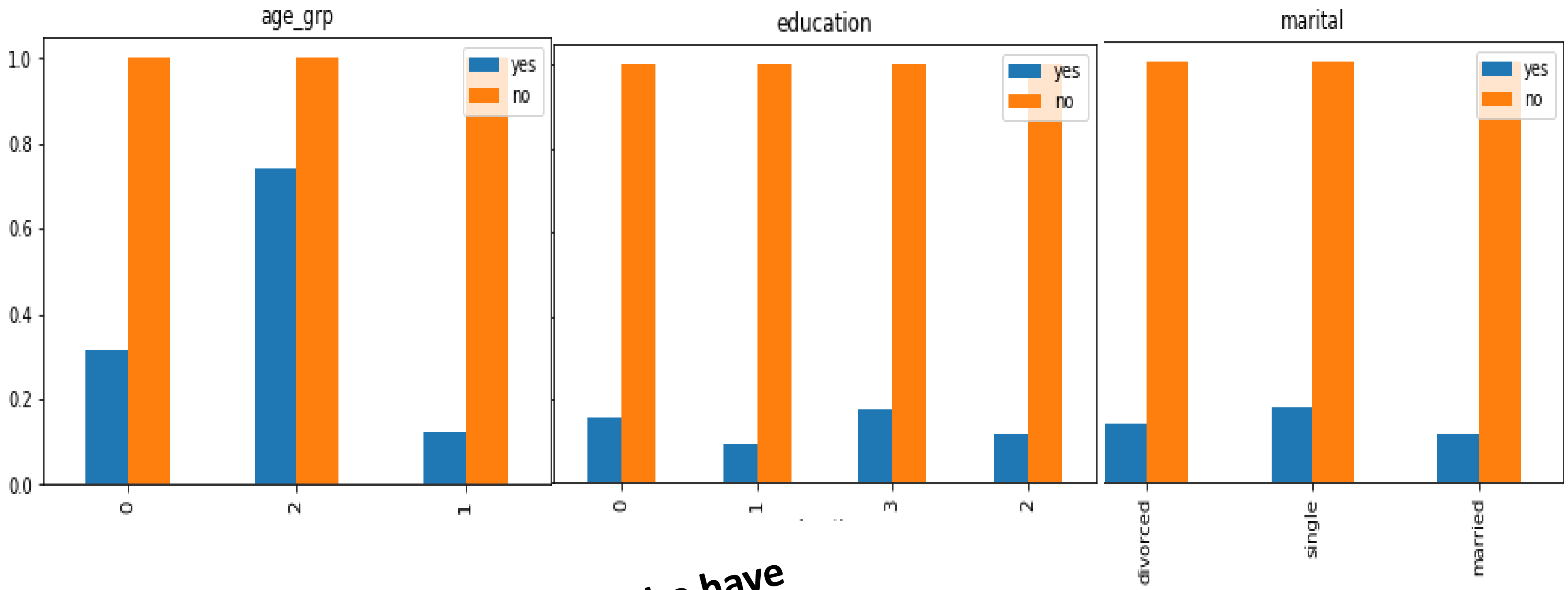
Data Analysis



Personal info & financial standing:
Who to target ?

Distribution of features

- age [18, 95]
- age_grp {0,1,2}
  - 0: young <25
  - 1: adult [25,65]
  - 2: elderly >65
- Education { 0,1,2,3}
  - 1: primary
  - 2: secondary
  - 3: tertiary
- Marital
  - Divorced
  - Single
  - Married

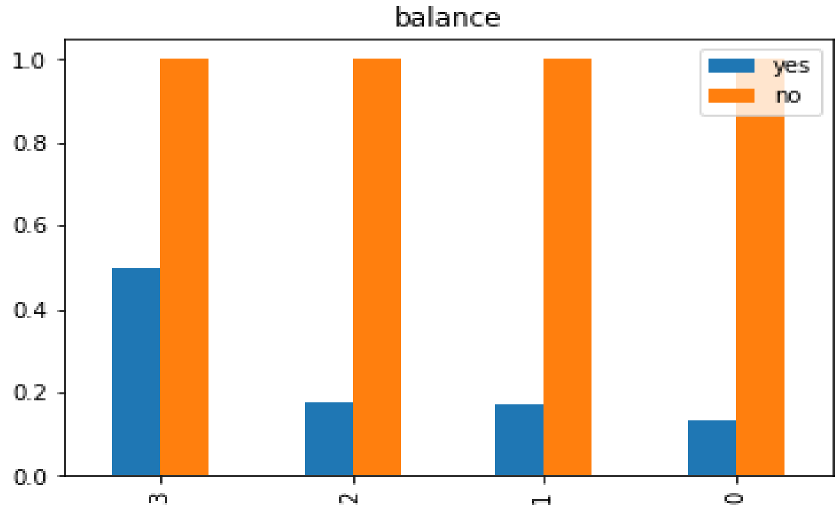


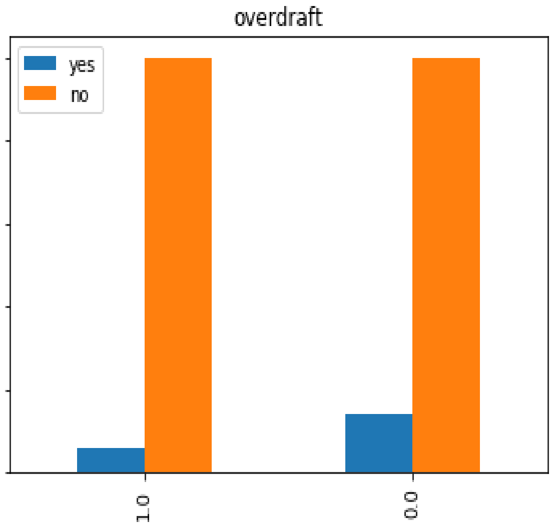
Single adults who have tertiarily education are a good starting point!

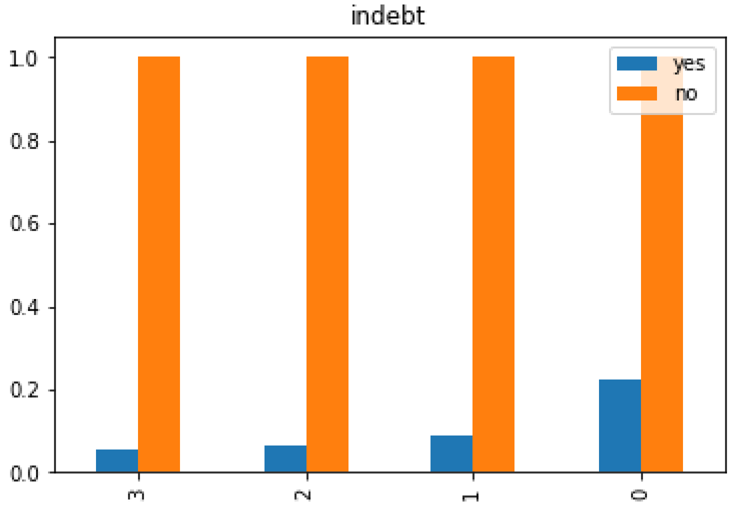
Personal info & financial standing:
Who to target ?

Distribution of features

- balance {0,1,2,3,4}
  - vlow, low, med, high, vhigh
- overdraft {0,1}
  - 1 if balance<0
- indebt: [0,3]
  - default+housing+loan
- default, housing, loan: {0,1}







+

Have high yearly balance, not indebt and with zero overdraft are a good starting point !

Personal info & financial standing:

Insights regarding client's lifestyle

- Elderly people have 0 overdraft

```
age_grp
0      112.0
1     3654.0
2        0.0
```

- Almost all elderlies (96%) have no debts

		housing
indebt	age_grp	
0	0	42.140719
	1	36.410863
	2	96.671105

- No defaults doesn't necessarily mean no overdrafts
- Costumers with 0 debts probably have no defaults

- Tertiarily educated adults have the highest balances

```
education  balance
0          0            4.113819
           1            1.886792
           2            5.000000
           3           16.666667
1          0           15.180257
           1            8.80503
           2            5.00000
           3           33.33333
2          0           51.41162
           1           28.93081
           2           30.00000
           3           33.33333
3          0           29.29429
           1           60.37735
           2           60.00000
           3           16.66666
           4          100.00000
Name: age, dtype: float64
```

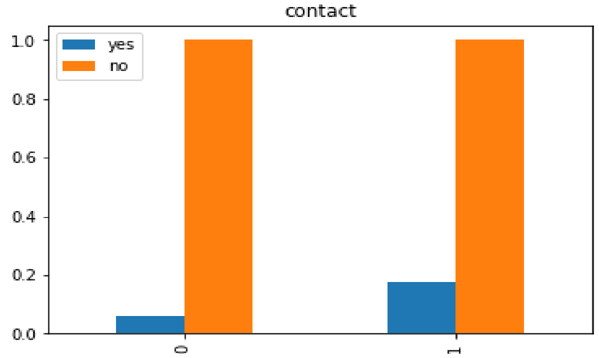
```
age_grp  balance
0        0            2.963193
         1            1.257862
1        0           95.390835
         1           93.710692
         2          100.000000
         3           66.666667
         4          100.000000
2        0            1.645972
         1            5.031447
         3           33.333333
Name: age, dtype: float64
```

Current campaign contact:
When and how to contact clients?

Distribution of features

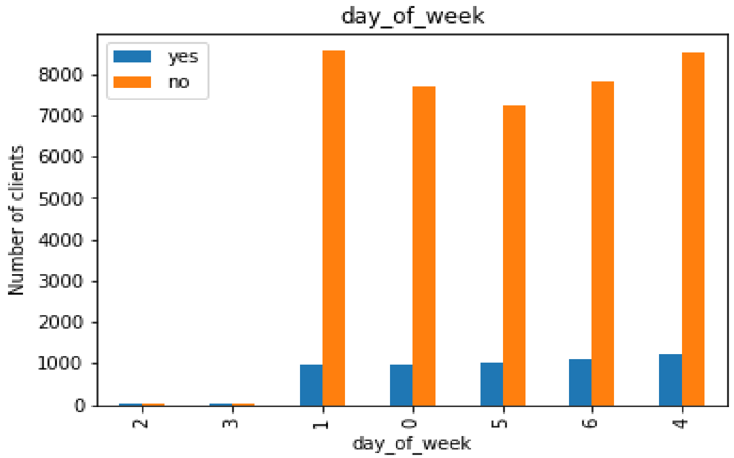
- year {2008,2009,2010}
- contact
  - 0: telephone
  - 1: cellphone
- campaign [1:36]
  - (#of contacts with 1 client)
- duration [1:81] (in minutes)
  - Last contact duration with 1 client
- month , day
- day_of_week [0:6]
  - Sunday→Saturday
- Rec {0:1}
  - 1 for 2008 and 2009 till June
  - 0 afterwards

- Cell contact is more effective than telephone and gained more popularity over years




```
year    contact
2008.0  0          50.863717
        1          49.136283
2009.0  0           8.769096
        1          91.230904
2010.0  0          19.824293
        1          80.175707
Name: contact, dtype: float64
```

- Bank employees focus their calls away from the middle of the week



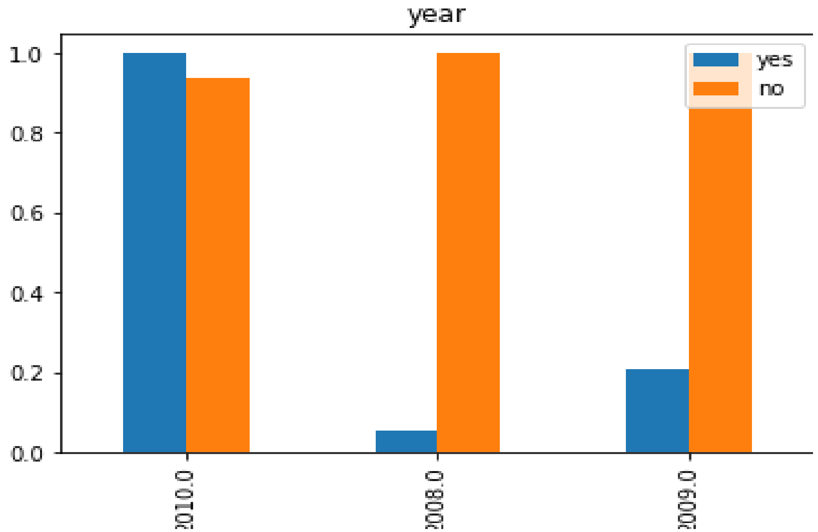
Current campaign contact:

Insights regarding factors affecting contact rate

- Highest number of contacted costumers is in 2008 (although we have 6 months record only)

And yet, very little people accepted the product!

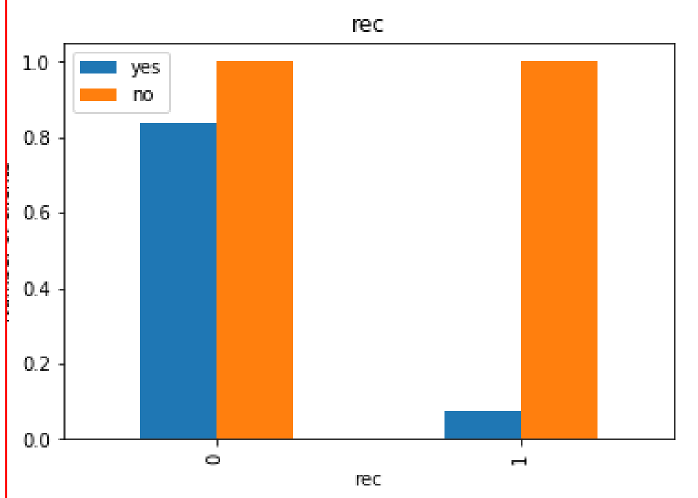
```
year
2008.0    27729
2009.0    14859
2010.0     2618
Name: age, dtype: int64
```



- The average duration and frequency of calls was highest in 2008

year	duration	campaign
2008.0	13	3.186447
2009.0	11	2.130291
2010.0	5	1.884263

The cause of these results is recession that started in 2008 and ended officially in June 2009



Previous contact with costumers

Features’ distribution and dependencies

- poutcome {0,1}
  - Outcome of previous outreach on past campaigns
- Previous [0,275]
  - #of contacts with this costumer before

- **When no previous contact exist, this means a costumer has not participated in any previous campaigns (poutcome: non-existent)**

So the cases for these features are:

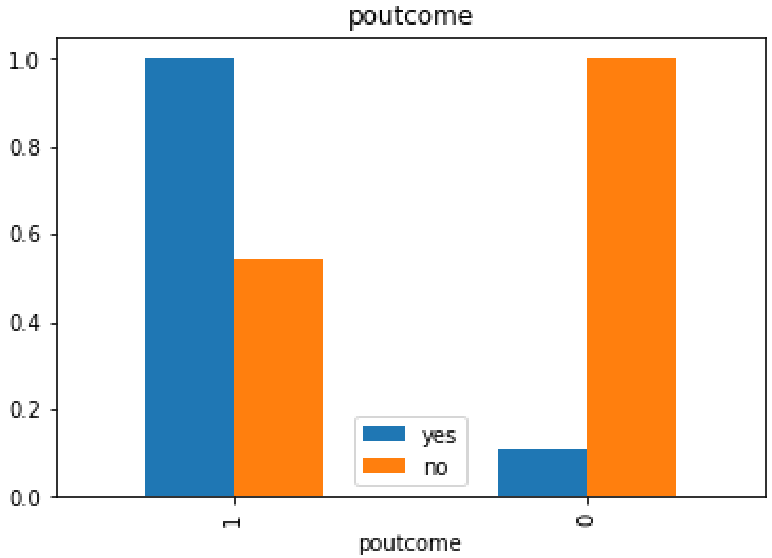
previous=0, poutcome=0 → non-existent

previous>0,poutcome= 0 or 1 → no or yes

Previous contact with costumers
Costumer's responsiveness to previous contact

- **Most dominant case in the dataset is the non-existent case**

```
# of entries where previous>0=8252
# of entries where previous=0=36954
```



Costumers who have previous contact with the bank are more inclined to say yes to current campaigns

Economic & social indicators

What can societal indicators tell us?

- Euribor3m: direct driver for setting interest rates on bank products

- Cosum_price_rate: reflects inflation

- Consum_conf_index: reflects economic growth

- Unemployed_rate

Proposed modeling techniques

Techniques

1) Logistic regression (benchmark of binary classification)
2) Random forest (resilient to ambiguous values such as “unknown”)
3) Balanced bagging classifier (re-sampling of majority and minority classes to combat imbalanced datasets)
4) XGBoost (hyperparameters are tuned to balance-out highly imbalanced datasets)

Performance criterion→ Percision

- Accuracy is misleading due to highly imbalanced datasets

- The metric is determined by deciding on the most critical type of error, in this project the biggest waste of resources is contacting clients who won’t purchase the deposit product → False positives

Thank You

Data Glacier

Your Deep Learning Partner